

The Scope of FBIS and BBC Open-Source Media Coverage, 1979–2008 (U)

Kalev Leetaru

For nearly 70 years, the Foreign Broadcast Information Service (FBIS) monitored the world's airwaves and other news outlets, transcribing and translating selected content into English and in the process creating a multi-million-page historical archive of the global news media. Yet, FBIS material has not been widely utilized in the academic content analysis community, perhaps because relatively little is known about the scope of the content that is digitally available to researchers in this field. This article, researched and written by a specialist in the field, contains a brief overview of the service—reestablished as the Open Source Center in 2004—and a statistical examination of the unclassified FBIS material produced from July 1993 through July 2004—a period during which FBIS produced and distributed CDs of its selected material. Examined are language preferences, distribution of monitored sources, and topical and geographic emphases. The author examines the output of a similar service provided by the British Broadcasting Corporation (BBC), known as the Summary of World Broadcasts (SWB). Its digital files permit the tracing of coverage trends from January 1979 through December 2008 and invite comparison with FBIS efforts.

“
Archival practices of usual news sources constrain scholarship, especially on cross national issues.
”

Social scientists rely heavily on archival news sources, but the selection and archival practices of these sources constrain scholarship, especially on cross-national issues. Contemporary news aggregators like Lexis Nexis focus on compiling large numbers of news sources into a single, searchable archive, but their historical files are limited. Historical sources like the Proquest Historical Newspapers Database offer news content back to the mid-19th century or earlier, but they include only a few publications. Both rely nearly exclusively on English-language Western news sources.

Global news databases like News-Bank's Access World

News primarily emphasize English-language “international” editions of major foreign newspapers, which often do not represent the views of a nation's vernacular news content. Nor do these services maintain the output of foreign broadcast media, especially critical in regions with low literacy rates. These limitations, for example, make it difficult to examine such questions as, how the international press cover the 2002 collapse of the American communications giant WorldCom or, in what ways did different regions of the world deal with the fallout and its impact on their domestic economies? Answering such questions on a truly international scale requires researchers to

All statements of fact, opinion, or analysis expressed in this article are those of the authors. Nothing in the article should be construed as asserting or implying US government endorsement of an article's factual statements and interpretations.

FBIS and BBC have served as strategic resources, maintaining relatively even monitoring volume across the globe on a broad range of topics.

have the ability to examine representative samples of news reports in countries from across the world—print, broadcast, and Internet.¹

The contents of FBIS and SWB collections currently available to academic researchers provide the material the commercial aggregators do not. During the period studied for this article (1993–2004 for FBIS and 1979–2008 for SWB) the services have served as strategic resources, maintaining relatively even monitoring volume across the globe on a broad range of topics, and thus provide an ideal foundation for cross-national content analysis.

In addition, the focus of the two services on broadcast material has offered critical visibility into regions where broadcasts are the predominant form of popular news distribution. The ability to select material by geographic and topical emphasis and to access English translations of vernacular content in print, broadcast, and Internet sources has made FBIS material, in particular, an unparalleled resource for content analysis of foreign media.

A Brief Historical Overview

Since the beginning of World War II, the United States and Great Britain have operated

the world's most extensive media monitoring services. Known eventually as open source intelligence (OSINT)—the collection and exploitation of noncovert information sources, including television and radio broadcasts, newspapers, trade publications, Internet Web sites, and nearly any other form of public dissemination. The two services have paid particular attention to vernacular-language sources aimed at domestic populations.

In some cases OSINT has been used simply to gauge local reaction to events. In other cases, it has been used to support estimates of future events or to identify rhetorical patterns or broadcast schedules to support intelligence analysis. One of the greatest benefits of OSINT over traditional covert intelligence has been its nearly real-time nature (material could be examined very soon after it was produced) and the relative ease and minimal risk of its acquisition and dissemination.

Newswire services like the Associated Press collect news from around the world, but they do so primarily through their own reporting staffs or stringers. A protest covered in a remote province of China is likely to be seen through the eyes of a Western-trained writer or photographer and

reflect Western views. A domestic broadcast or newspaper article, on the other hand, reflects the perspectives of local populations or local authorities, depending on the degree of government control of the media, both in its factual reporting and the words used to convey that information. The global news media form a very nonhomogeneous distribution layer and news outlets are subject to strong cultural and contextual influences that may be explored through the ways in which they cover events.²

Known affectionately as “America’s window on the world,”³ FBIS was the backbone of OSINT collection in the US Intelligence Community (IC), acting as the US government’s primary instrument for collecting, translating, and disseminating open-source information. FBIS analysts also played primary roles in analyzing open source information and synthesized large amounts of material into targeted reports. The importance of FBIS to the modern intelligence world was summed up in a *Washington Times* article in 2001: “so much of what the CIA learns is collected from newspaper clippings that the director of the agency ought to be called the Pastemaster General.”⁴

Wartime

The roots of institutionalized OSINT collection in the United States can be traced back to the Princeton Listening Center located in the Princeton Univer-

sity School of Public and International Affairs. Funded by the Rockefeller Foundation⁵ the center began operations in November 1939 with a mission to “monitor, transcribe, translate, and analyze shortwave propaganda broadcast[s] from Berlin, London, Paris, Rome, and, to some extent, Moscow.”⁶

A wide range of radio products was monitored, including “news bulletins, weekly topical talks, radio news reels, features and dramatizations.” Its limited staff could only record and analyze a sampling of broadcasts for propaganda content. Topics covered included how “propaganda varied between countries, as well as from one show to another within the same country ... the way in which specific incidents were reported, atrocity references, attitudes toward various countries, and the way this propaganda affected US listeners.” By April 1941, the listening center had compiled over 15 million words of transcribed material from English, German, French, and Italian short-wave broadcasts.

On 26 February 1941, President Roosevelt established the Foreign Broadcast Monitoring Service (FBMS) with orders to monitor foreign shortwave radio broadcasts from “belligerent, occupied, and neutral countries” directed at the United States.⁷ FBMS transcribed these broadcasts and used them to perform “trend analysis to discover shifts in tenor or con-

Intelligence gathering in the uncertain post-WW II world required sweeping up a wider range of international media broadcasts—too great a task for FBIS to realistically take on by itself.

tent that might imply changes in Japanese intentions.” The Princeton Listening Center became the core of the new agency and by the end of 1942, it was translating 500,000 words a day from 25 broadcasting stations in 15 languages.⁸ FBMS published its first transcription report on 18 November 1941 and its very first analytical report, dated 6 December 1941, contained the poignant statement:

*Japanese radio intensifies still further its defiant, hostile tone; in contrast to its behavior during earlier periods of Pacific tension, Radio Tokyo makes no peace appeals. Comment on the United States is bitter and increased; it is broadcast not only to this country, but to Latin America and Southeastern Asia.*⁹

The Cold War

On 15 August 1945 FBIS recorded Emperor Hirohito’s surrender announcement to the Japanese people, and on 14 December it published its final wartime daily report, having proved its utility to intelligence during the war. With the approbation of the *Washington Post*, which called the service “one of the most vital units in a sound postwar intelligence operation,” the service was transferred to the Central Intelligence Group of the National

Intelligence Authority, forerunners of the CIA.¹⁰

Wartime intelligence gathering required significant resources, but they could be directed toward a small number of countries and sources. Intelligence gathering in the uncertain post-WW II world required sweeping up a wider range of international media broadcasts—too great a task for FBIS to realistically take on by itself. Fortunately for Allied postwar intelligence, the United Kingdom had developed its own open source intelligence service, the British Broadcast Monitoring Service, just prior to the war. From its founding on 22 August 1939, it produced a foreign broadcast compilation called the *Digest of World Broadcasts*—renamed the *Summary of World Broadcasts* in May 1947.¹¹

By 1945, the BBC service was monitoring 1.25 million words per day in 30 languages, although limited resources allowed translation into English of only 300,000. FBIS, on the other hand, transcribed and translated the majority of the content it monitored.¹² Subsequently, a British-US agreement led to a cooperative media coverage and sharing arrangement that has lasted to the present day.¹³ As a result of the agreement, BBC has generally focused on Central Asia and

The power of OSINT to peer into closed societies, to predict major events, and to offer real-time updates cannot be overstated.

nations that were part of the Soviet Union; FBIS has handled the Far East and Latin America, and the two services jointly have covered Africa, the Middle East, and Europe. The agencies also agreed to operate under similar “operational and editorial standards.”¹⁴

Radio broadcasts and press agency transmissions were the original focus of FBIS, which added television coverage as it became more popular. Print material became a focus of FBIS only in 1967, and by 1992, its mission had expanded to include commercial and governmental public-access databases, and gray literature (“private or public symposia proceedings and academic studies”).¹⁵ Even though it did not adopt print material until 1967, substantial news reports were usually carried by press agencies on their wirefeeds, which FBIS monitored nearly from the beginning. By 1992, the service had developed a network of 19 regional bureaus, which served as collection, processing, and distribution points for their geographic areas.¹⁶

FBIS and BBC have emphasized historically reliable or authoritative sources, but FBIS continually adds new sources and a “not insignificant amount of [its] total effort is spent identifying and assessing sources to ensure the reliability, accuracy,

responsiveness, and completeness of ... coverage.”¹⁷ By 1992, FBIS was monitoring more than 3,500 publications in 55 languages and 790 hours of television a week in 29 languages from 50 countries.¹⁸

The power of OSINT to peer into closed societies, to predict major events and to offer real-time updates cannot be overstated. Its utility in the intelligence analysis process has been the subject of numerous studies and the testimony of any number of senior intelligence officials. Suffice it to say here that former Deputy Director of Central Intelligence William Studeman estimated in a 1992 speech frequently cited in this essay that more than 80 percent of many intelligence needs could be met through open sources.¹⁹ By the late 1990s, FBIS was serving much more than IC needs: a 1997 study showed that the Law Library of Congress was relying heavily on FBIS to provide “quality and [timely] information to Congress about legal, legal-political and legal-economic developments abroad.”²⁰

The “basket of sources” nature of OSINT has allowed it to leverage the combined reporting power of multiple sources, reaching beyond the limitations of any single source. A 2006 study examining the use of OSINT material for event

identification from news material found the *Summary of World Broadcasts* to be dramatically superior in volume and breadth to traditional commercial newswires.²¹ Newswires, with their larger reporting infrastructure and geographic coverage than newspapers, still rely on a single set of reporters to cover every country. OSINT compilations like FBIS and SWB, on the other hand, repackaged content from across the entire globe, combining the viewpoints of multiple outlets while maintaining fairly comprehensive coverage of national presses.²²

Having briefly, in 1996, faced extinction, FBIS was reborn in the wake of the Intelligence Reform and Terrorism Protection Act of 2004 as the Open Source Center under a newly created Office of the Director of National Intelligence. In his remarks at a ceremony marking OSC’s creation, General Michael Hayden, then the deputy director of national intelligence, noted that OSC “will advance the Intelligence Community’s exploitation of openly available information to include the Internet, databases, press, radio, television, video, geospatial data, photos and commercial imagery.”²³

By 2006, OSC reportedly had “stepped up data collection and analysis to include bloggers worldwide and [was] developing new methods to gauge the reliability of the content.” The report noted that in order to

expand OSINT efforts, OSC had doubled its staff and become a clearing house for material from 32 different US government OSINT units, and its translators turned more than 30 million words a month into English from languages across the world.²⁴

FBIS as the Public's Open Source

Designed to provide the Allies an advantage during WW II, FBIS, and its successor, has the added potential to be a critical resource for academic scholars, yet the scholarly community's lack of familiarity with open source methods and the FBIS collection in particular, has limited academic use of the FBIS archive. That archive already includes some of the material mentioned in Hayden's speech—print, broadcast, and Internet-derived material—translated into English and tagged by country and topic and is an unparalleled resource for understanding news content throughout the world across the last half-century.

FBIS reports became widely available for public use, in print and microfiche forms, in 1974, when the Commerce Department's National Technical Information Service (NTIS) began commercial distribution of the material.²⁵ In his 1992 speech, Admiral Studeman indicated a strong appreciation of the private-sector and academic research that had arisen

[FBIS] customers ... value the work of private sector scholars and analysts who avail themselves of our material and contribute to the national debate on contemporary issues.

out of FBIS's availability outside the US government and expressed a commitment to its continued availability. As he noted, "FBIS's customers in both the intelligence and policy communities ... value the work of private-sector scholars and analysts who avail themselves of our material and contribute significantly to the national debate on contemporary issues."²⁶

The following year, 1993, FBIS began to distribute CDs of its material to Federal Depository Libraries, a practice that lasted until June 2004, when FBIS began Internet-only distribution through Dialog Corporation's World News Connection (WNC) service (<http://wnc.fed-world.gov/>), which licenses the material from the US government. This Web-based portal offers hourly updates and full text keyword searching of FBIS material from January 1996 to the present.

The CD collection allows greater flexibility in accessing reports than the Dialog interface. Dialog only displays 10 results at a time and offers limited interactive refinement capabilities. The inaugural CD issued in 1993 covers a period of nearly one year, but only a small number of reports are included for the period November 1992–June 1993. July–September 1993 is fully covered.

Thereafter, into June 2004, each distributed CD covered periods of three months.



The FBIS Dashboard

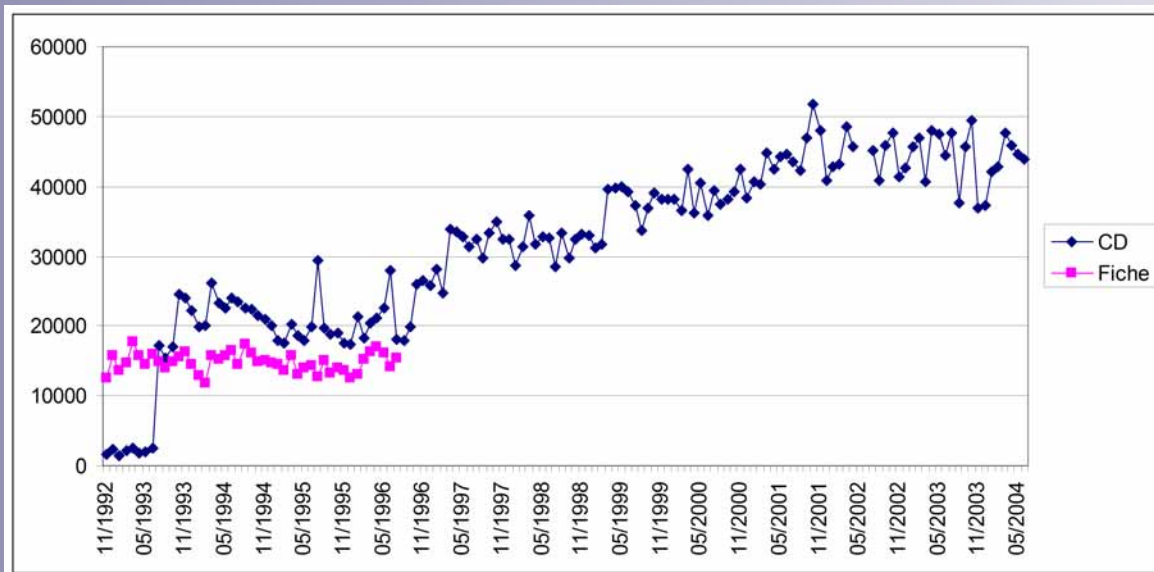
The Pulse of Activity

FBIS collection during the decade following the end of the Cold War, as seen in figure 1, reflects a relatively stable monthly volume through the end of 1996, when growth started climbing steadily into early 2001, when it stabilized again. As noted above, FBIS faced severe cuts in 1996, before an outpouring of public support contributed to its survival. This graph indicates that the service not only survived but found ways (and resources) to allow it to more than double its monthly output during in the next five years.

The Nature of the Material

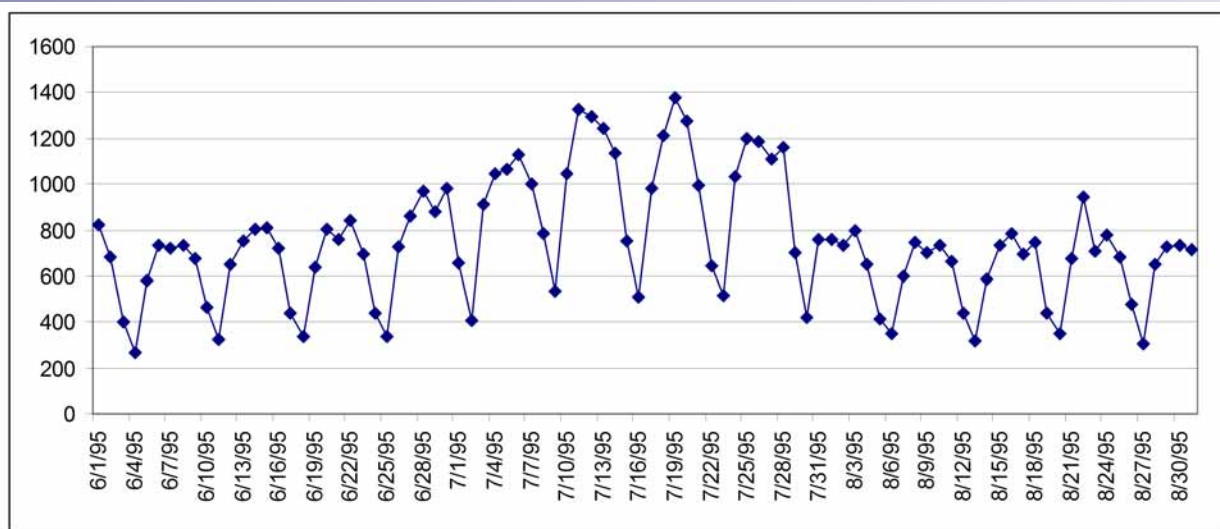
While its primary focus is on news material, FBIS also captures editorial content and commentaries, which its monitors tag at the beginning of reports. Such reports constitute 6.3 percent of the collection—3.5 percent are flagged as editorial content and 2.8 percent as commentaries. Editorial and commentary content represented 5–6 percent of each year's total reports through 1999, but in 2000 the percentage increased nearly 1 percent each year to a

Figure 1: Monthly FBIS Volume, November 1992–June 2004



During this period, FBIS compiled 4,393,121 reports. The monthly distribution of these reports as collected in the CDs is shown in blue. The low number in the first months reflects the small number of reports transferred to CD at the beginning of the effort. The magenta points show the number of titles listed in an index of printed FBIS reports prepared under contract by NewsBank, Inc. Newsbank's index shows a lower volume of reports (about 30 percent less on average per month), possibly because apparent duplicate reports were not listed. (No copy of CD #39 (May/June 2002) could be located and could not be included in this analysis.)

Figure 2: Daily FBIS Volume, June 1995–August 1995.



Daily reporting volumes, as seen in this three-month snapshot from 1995, indicate that FBIS daily reporting patterns resemble those of major news aggregators, except that FBIS' lowest volumes occur on Sundays instead of Saturdays. This may reflect FBIS staffing patterns or other factors in international news activity.

Table 1: Top 25 Source Languages

Origin Language	Report Count	% All Reports
English	2021021	46.00
Russian	371106	8.45
Arabic	271326	6.18
Spanish	197451	4.49
French	138046	3.14
Serbo-Croatian	135805	3.09
Chinese	124014	2.82
Persian	80720	1.84
German	76688	1.75
Portuguese	66003	1.50
Turkish	65951	1.50
Hebrew	51670	1.18
Japanese	50509	1.15
Korean	47113	1.07
Albanian	40898	0.93
Italian	39060	0.89
Urdu	31705	0.72
Ukrainian	31608	0.72
Indonesian	29359	0.67
Greek	28564	0.65
Polish	28372	0.65
Hungarian	26392	0.60
Slovak	22980	0.52
Bulgarian	22920	0.52

Table 2: Topics Covered, 1999–2004

Topic	Report Count	% Reports
Domestic Political	1204515	44.94
International Political	1164586	43.45
Leader	927241	34.59
Military	459898	17.16
Domestic Economic	411593	15.36
International Economic	357610	13.34
Terrorism	277667	10.36
Urgent	245054	9.14
Human Rights	196205	7.32
Political	187128	6.98
Crime	129829	4.84
International	128016	4.78
Domestic	116710	4.35
Dissent	101710	3.79
Media	84157	3.14
Energy	83920	3.13
Technology	63072	2.35
Proliferation	63003	2.35
Peacekeeping	59076	2.20
Environment	55814	2.08
Economic	55157	2.06
Health	49847	1.86
Migration	40435	1.51
Telecom	37711	1.41
Narcotics	35017	1.31
Conflict	32656	1.22

Table 3: Top 25 Media Outlets

Source	Report Count	% All Reports
Beijing XINHUA	194316	4.42
Moscow ITAR-TASS	155925	3.55
Tokyo KYODO	123404	2.81
Seoul YONHAP	92722	2.11
Tehran IRNA	57857	1.32
Paris AFP	56286	1.28
Hong Kong AFP	44390	1.01
Prague CTK	39201	0.89
Ankara Anatolia	31436	0.72
P'yongyang KCNA	29824	0.68
Moscow INTERFAX	29141	0.66
Belgrade BETA	28717	0.65
Belgrade TANJUG	28381	0.65
Cairo MENA	26764	0.61
Pyongyang KCNA	26230	0.60
Zagreb HINA	23013	0.52
Taipei Central News Agency WWW-Text	22983	0.52
Moscow RIA	22071	0.50
Tokyo Jiji Press	21508	0.49
Moscow Nezavisimaya Gazeta	20371	0.46
Moscow Agentstvo Voenyenykh Novostey WWW-Text	19931	0.45
Jerusalem Qol Yisra'el	19896	0.45
Madrid EFE	18973	0.43
Warsaw PAP	17903	0.41

peak of just over 9 percent in 2003.

The proportion of excerpted reports over the study period was relatively low, —averaging around 5.6 percent per year— making FBIS material ideal for content analysis. Longer broadcast or print reports are excerpted when only portions of an item are relevant to targeted subject areas. For example, a Radio France International broadcast might have been excerpted to transcribe just those comments about an African country's denunciation of a trade embargo against it or a brief mention of a party official's death in a People's Republic of China radio broadcast might be extracted from other unimportant material.²⁷

Language

English-language material comprises approximately 46

percent of the material FBIS collects. Such material represents a saving in translation expenses and, when coming from media controlled by authoritarian regimes, potentially authoritative messages to US and other Western governments. Table 1 shows the top 25 source languages for FBIS reports during 1992–2004. After English, Russian and Arabic reports were the most frequently collected.

Topics

On 1 January 1999, FBIS began to include topical category tags in its reports, each of which could have as many tags as necessary to fully describe its contents. As table 2 shows, however, political issues topped FBIS collection, comprising nearly 83 percent of all content. Economic issues accounted for 26 percent. From January to July 1999, reports were also categorized sepa-

rately as “international” or “domestic” and “political” or “economic.” In August 1999 the specialized categories “domestic political,” “international political,” “domestic economic,” and “international economic” were introduced. All other categories ran continuously from January 1999 until the end of this sampling period.

Media Outlets

Content analysts must consider the volume of material produced by each source to ensure that no one media outlet dominates in their analyses. Table 3 lists the top 25 media outlets from which FBIS selected content during the study period from a universe exceeding 32,000 sources. (Because FBIS citations often distinguish between Web and print editions of a source and between different editions of a source—international, regional, local, weekend editions—the

Understanding the physical location of each source is critical to exploring possible geographic biases in monitoring.

actual number of unique sources noted in the table is probably significantly lower than the number shown.) In any case, taken together, selections from the top 25 outlets accounted for more than a quarter of all FBIS-selected material during this period. Though this small proportion of the world's media outlets dominated FBIS collection, they are outlets with national stature and international importance.

The Geography of Coverage

Understanding the physical location of each source is critical to exploring possible geographic biases in monitoring. Unfortunately, while FBIS source references do indicate the geographic location of sources, they do not do so in a regular format, so an extensive machine geocoding system was used to automatically extract and compute GIS-compatible latitude and longitude coordinates for each FBIS source. In all, coordinates were calculated for 97.5 percent of reports and a random sample of 100 entries checked by hand showed no errors.

The maps on the following pages (figures 3–6) subdivide sources by geographic location, situating each in its listed city of origin. Immediately noticeable are the strong similarities between the maps, showing that FBIS heavily overlapped its coverage in each region,

combining broadcast, print, and Internet sources together. This mitigated the potential biases of any one distribution format. For example, in the Arab media, low general literacy rates mean that broadcast media formed the primary distribution channel for the masses and so is subjected to greater censorship than print media, which targets the elite.²⁸

After print material was added to FBIS collection in 1967, it became the dominant source for FBIS reports, constituting just over one-half of FBIS sources during the study period. (See figure 4.) To determine the source type of each outlet, the full reference field of each report was examined. Any reference that contained a time stamp (such as 1130 GMT) was considered a broadcast source, while those containing the keywords “Internet,” “electronic,” or “www” were flagged as Internet editions. All remaining sources were assumed to be print sources.

As table 3 illustrates, some sources contributed a much larger volume than others, so the total number of reports gathered from sources of each type was also computed. A total of 25 percent of reports were from print sources, 25 percent were from Internet sources, and 51 percent were from broadcast sources. (See figures 5 and 6.) Thus, more than half of all

reports during the study period were attributed to broadcast outlets, in keeping with the FBIS broadcast heritage. This also makes conceptual sense in that broadcast outlets traditionally operate 24/7, while print outlets usually issue only a single edition each day, meaning there is far more broadcast material to monitor. A smaller number of broadcast stations transmitting throughout the day will be able to generate far more content than a large number of print outlets with a limited amount of page space.

Figure 7 shows the geographic distribution, by country, of monitored reports. It is important to note that developed countries (for example, France) may act as reporting surrogates for lesser developed neighbors or for countries in which their sources have interest. The sources in the developed countries, of course, also have better established media distribution networks. Since there is no independent, authoritative master list of media outlets by country that covers print, broadcast, and Internet sources, there is no way of knowing what percentage of the media in each country and the total news volume they generated was captured by FBIS.

In January 1994, FBIS editors began assigning geographic tags to their reports. Geographic tags describe the geographic focus of a report—not the location of a report's

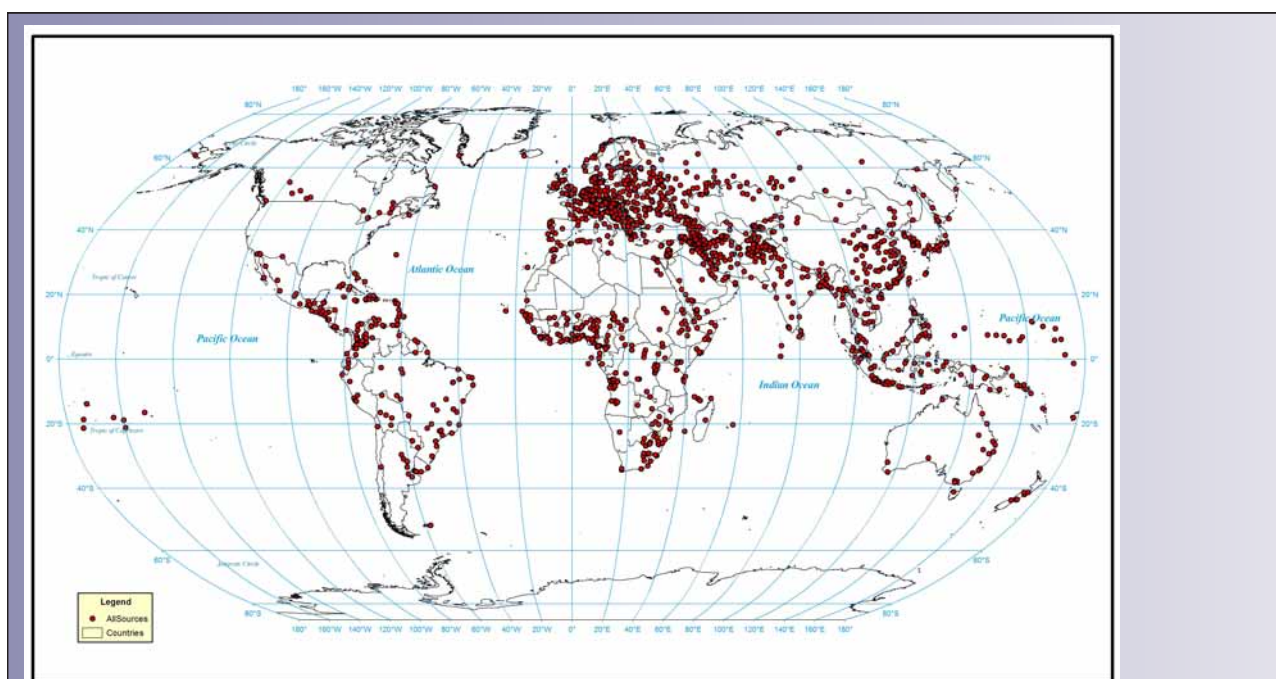


Figure 3: Locations of all sources monitored by FBIS during 1992–2004. About 83 percent of the shown locations are national capitals.

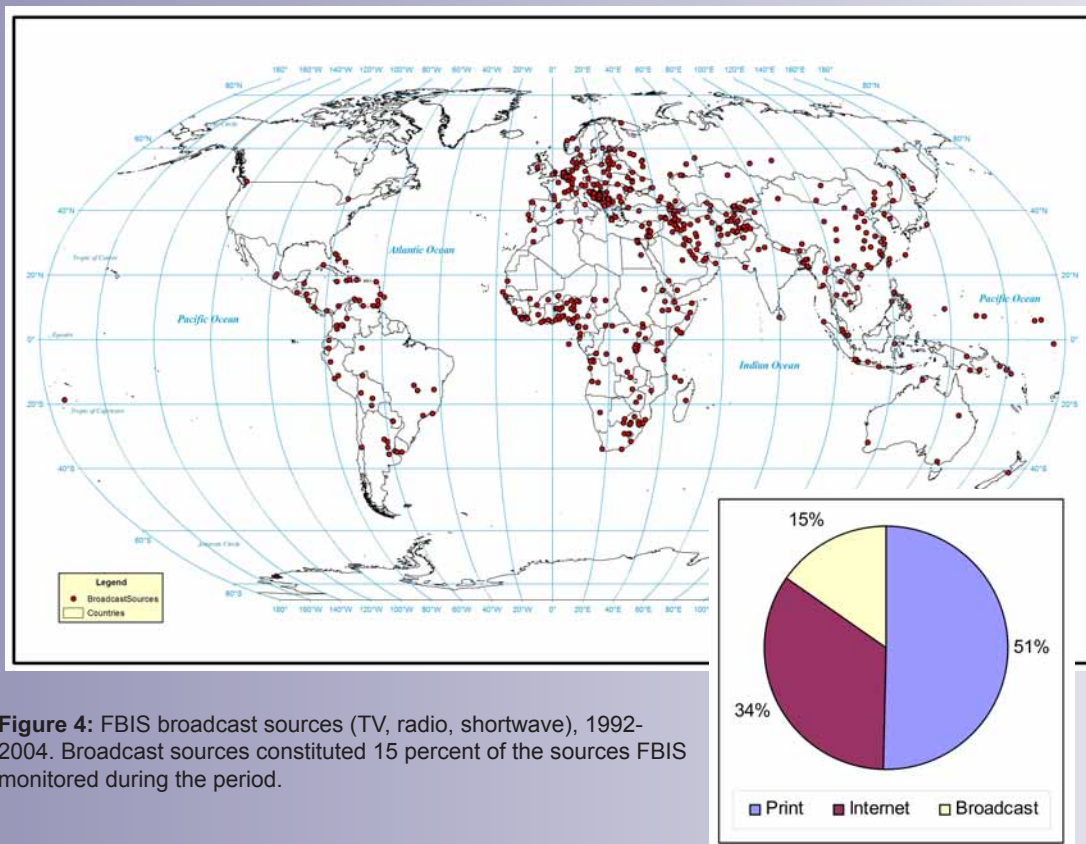


Figure 4: FBIS broadcast sources (TV, radio, shortwave), 1992–2004. Broadcast sources constituted 15 percent of the sources FBIS monitored during the period.

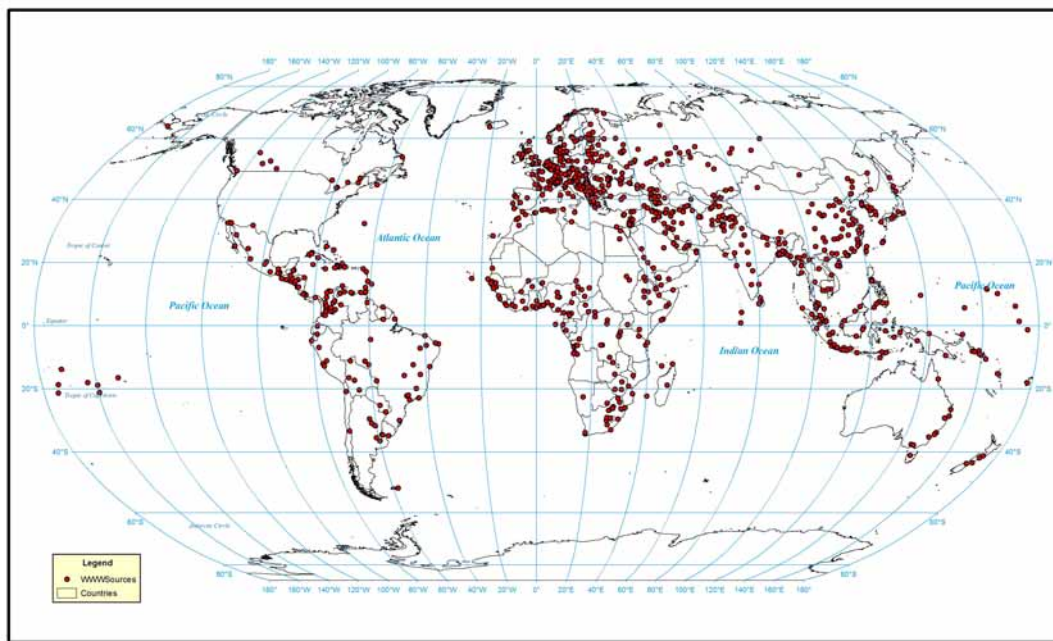


Figure 5: FBIS Internet source locations. These include Internet-only and Internet editions of print sources monitored during 1992–2004.

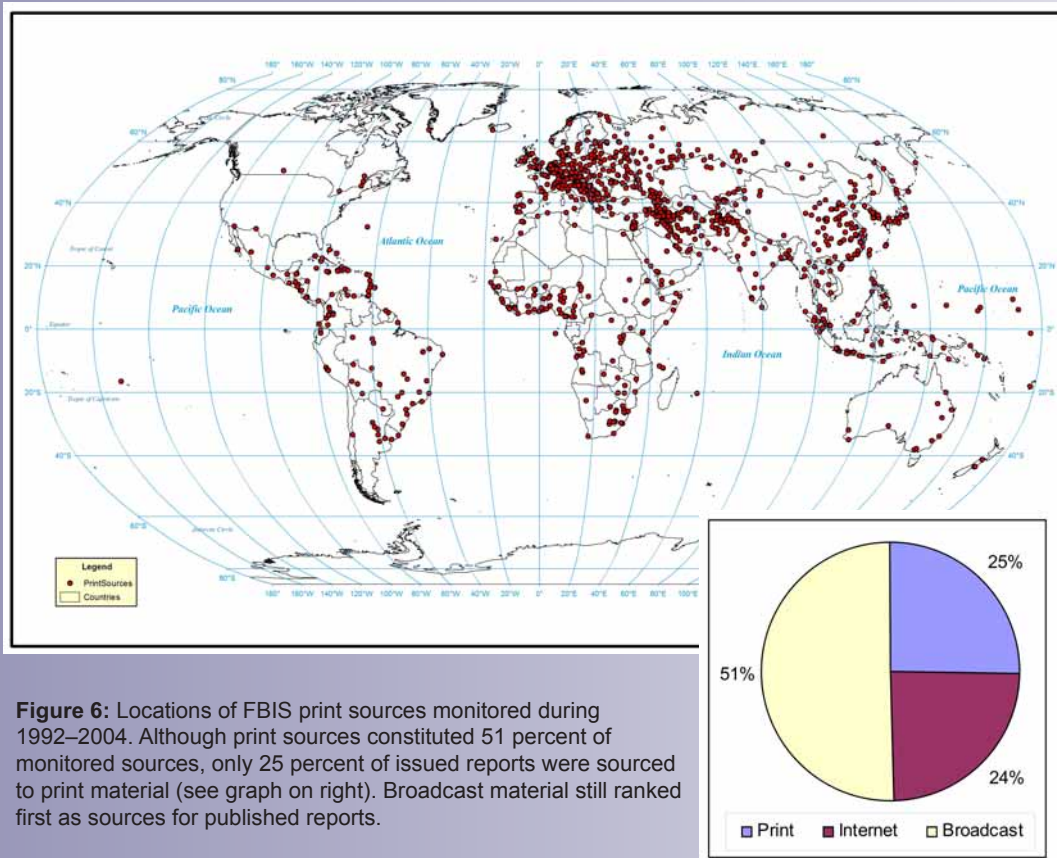


Figure 6: Locations of FBIS print sources monitored during 1992–2004. Although print sources constituted 51 percent of monitored sources, only 25 percent of issued reports were sourced to print material (see graph on right). Broadcast material still ranked first as sources for published reports.

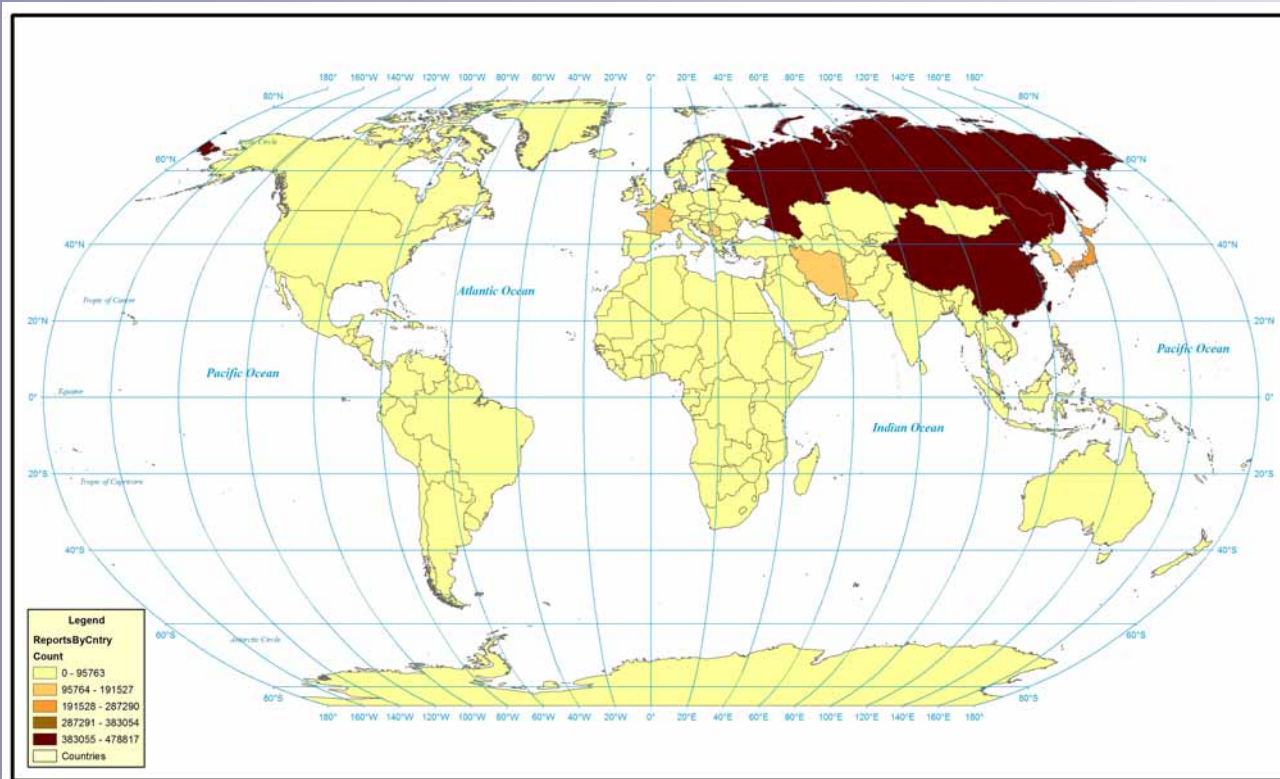


Figure 7: Reports by source country, 1992–2004.

source. A Chinese newspaper article describing events in India would have a tag only for India and not China, unless China played a major role in the report's contents. Combining the geographic information from the source reference with the geographic tags makes it possible to search for reports from one country that describe events in another country. Despite the potential for bias toward activities related to the United States, only 12 percent of articles published during this period actually had geographic tags for the United States, although the United States is the most frequently applied tag. (See table 5.) During this

period, Russia was the second most frequently tagged country.

A critical question in the study of this material is whether there has been any systematic bias toward monitoring a greater number of sources or gathering a greater number of reports in countries deemed to be hot spots by the United States. Alternatively, FBIS might have gathered reports uniformly across the world but focused primarily on those about the United States. Figure 7 shows that China and Russia provided the most material, more than 20 percent of all reports in the CDs from this period. Together with the

Table 4: Top 25 countries by number of articles from sources in that country, 1994–2004

Country	Report Count	% Reports
Russia	478817	10.90
China	466682	10.62
Japan	216446	4.93
Iran	170214	3.87
South Korea	152083	3.46
France	145677	3.32
Serbia & Montenegro	143009	3.26
United Kingdom	93609	2.13
Turkey	81982	1.87
North Korea	78845	1.79
India	74019	1.68
Belgium	69070	1.57
Germany	66887	1.52
Israel	62254	1.42
Bangladesh	60994	1.39
Egypt	59810	1.36
Bosnia & Herzegovina	58579	1.33
South Africa	58273	1.33
Czech Republic	58209	1.33
Italy	54091	1.23
Bulgaria	47388	1.08
Indonesia	45243	1.03
Ukraine	43632	0.99
Romania	43015	0.98
Poland	42309	0.96

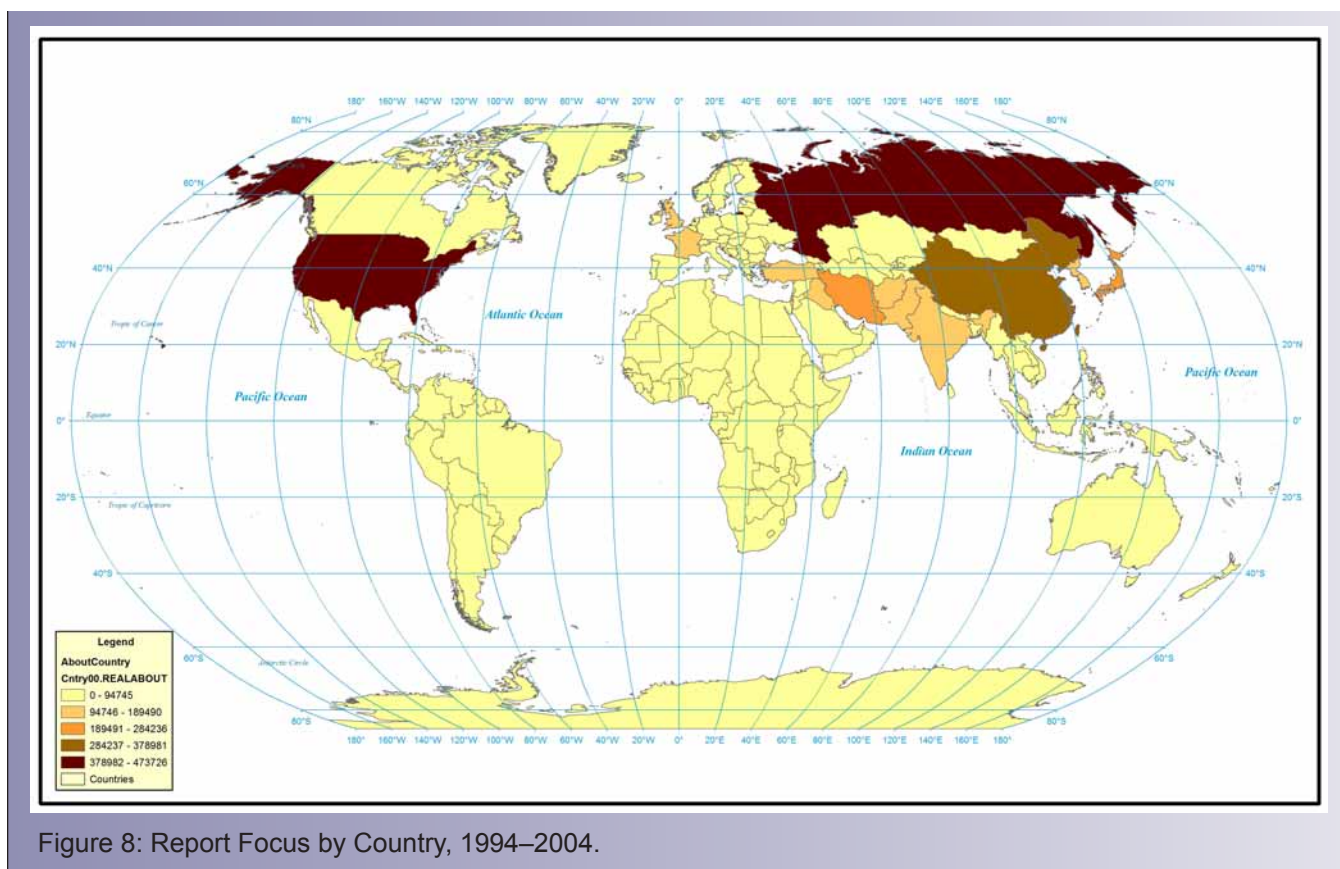


Table 5: Top 25 countries mentioned in all reporting, 1994–2004

Country	Report Count	% Reports
United States	473726	12.20
Russia	420446	10.83
China	337852	8.70
Japan	247578	6.38
Iran	193618	4.99
Israel	188803	4.86
South Korea	179029	4.61
Iraq	173234	4.46
North Korea	138658	3.57
India	134086	3.45
Pakistan	131563	3.39
United Kingdom	124512	3.21
Turkey	114355	2.95
West Bank & Gaza Strip	110829	2.86
Afghanistan	102059	2.63
France	96178	2.48
Germany	89763	2.31
Federal Republic of Yugoslavia	82788	2.13
Taiwan	82115	2.12
Serbia	78144	2.01
Kosovo	75940	1.96
Egypt	71914	1.85
Bosnia Herzegovina	70565	1.82
Italy	63011	1.62
Indonesia	58845	1.52

United States, China and Russia account for more than 30 percent of the geographic focus of all reports. (See figure 8.) However, Russia and China are also regional superpowers having significant interaction with their neighbors in the Eastern Hemisphere and thus are ideally positioned to report on events in that region.

Since reports collected in a given country are not necessarily about that country, useful is a comparison of the percentage of all reports sourced from a country with those having a geographic topic tag for that country. Figure 9 shows geographic sources and sinks—countries (in blue) about which

more reports are collected from outside their borders than from within their borders. South America is net neutral overall, with similar volumes of reports being sourced from each country as are monitored and reported about that country.

Africa as a whole is a net sink, with many more reports produced about that continent than are sourced from it. This is both the result of relatively underdeveloped media distribution networks and greater barriers to collection of material from African locations. This reality presents significant challenges to analysts, who must deal with content about these nations collected from

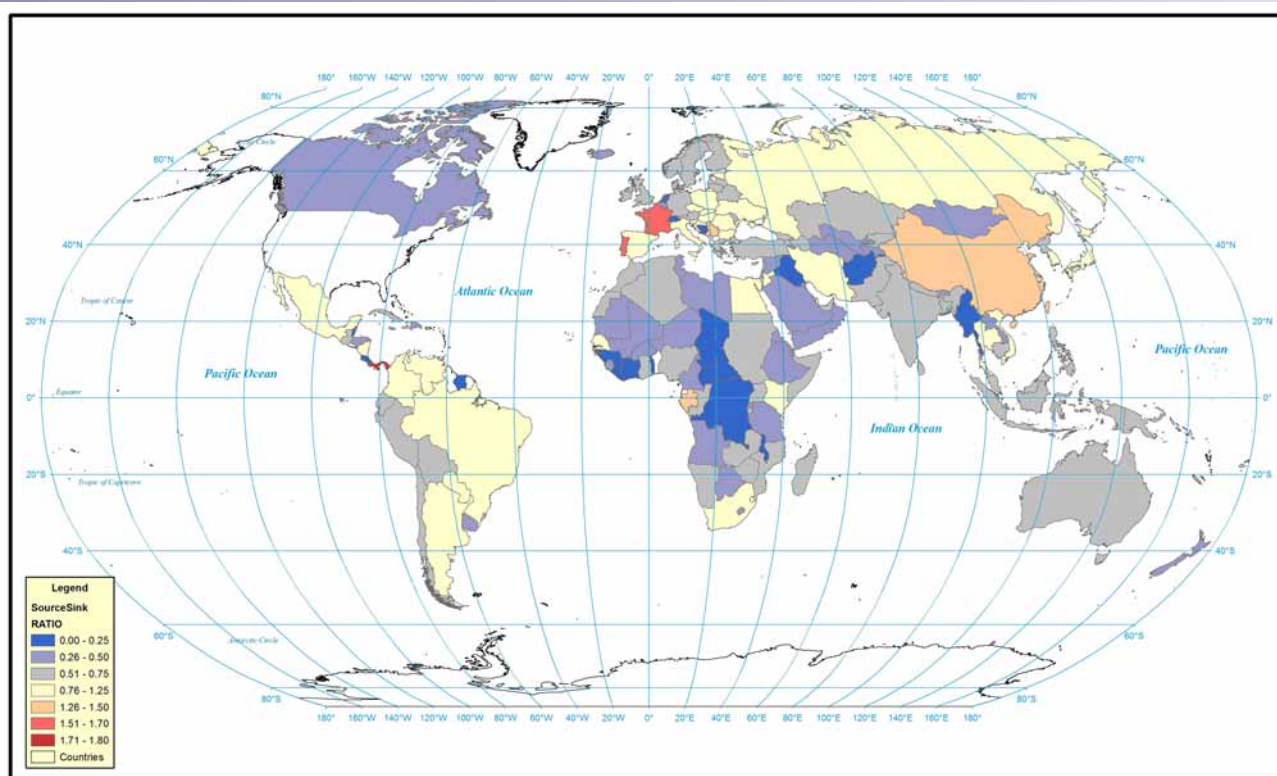


Figure 9: FBIS source (orange to red) and “Sink” (gray to blue) countries, 1994–2004.

outside their borders and subject to foreign, rather than domestic, views on internal events. By contrast, France is a net source, largely because of the presence of Agence France Presse (AFP) wire service. Similarly, BETA and TANJUG news agencies in Belgrade contributed to Serbia’s ranking as a net source during this period.

The coverage statistics do not appear to indicate that FBIS appreciably favored regions in which the United States was actively engaged during 1994–2004. The figures reflect a fairly even coverage outside Russia and China without redirecting resources toward more prob-

lematic regions. This suggests that FBIS provided a strategic service, monitoring all regions of the world relatively evenly rather than a tactical resource focused on troublesome areas. This is a critical attribute for using this material in content analysis.

❖ ❖ ❖

BBC Summary of World Broadcasts

Whereas public access to historical digital FBIS content only began in July 1993, and public access to content after 2004 is limited by the technical constraints of the Dialog search interface, material from the SWB service has been available since 1 January 1979 through LexisNexis. Like FBIS, SWB today monitors media from 150 countries in more than 100 languages from over 3,000 sources. It has overseas bureaus in Azerbaijan, Egypt, India, Kenya, Russia, Ukraine, and Uzbekistan and a staff of around 500.²⁹ It has a wide cor-

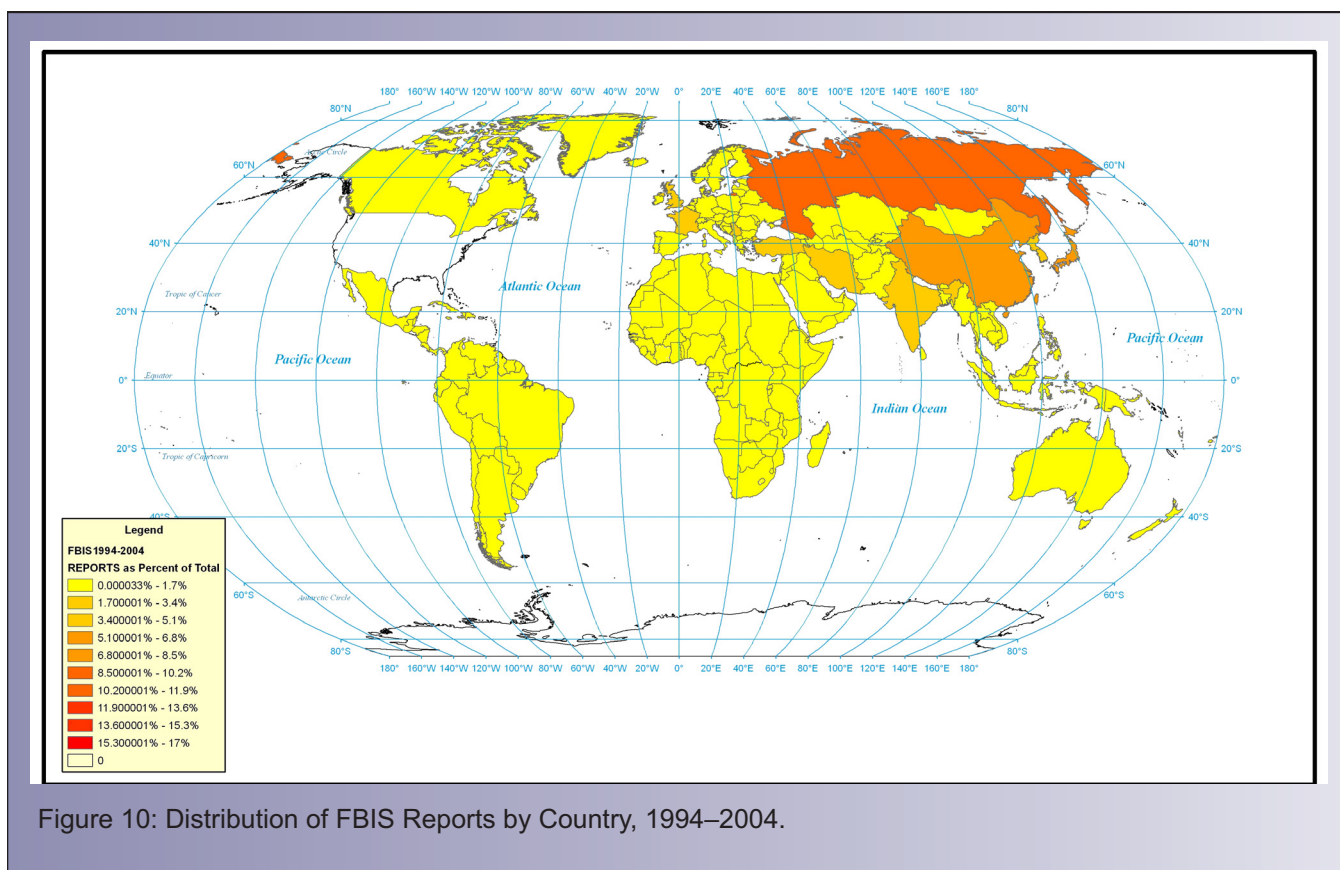


Figure 10: Distribution of FBIS Reports by Country, 1994–2004.

porate following, first appearing in the Reuters Business Briefing newswire in 1983, and in 2001 was one of the 10 most popular news sources in that service.³⁰

SWB’s mission is to focus on “political, economic, security, and media news, comment, and reaction.” The service acknowledges geographic prioritization: Iraq and Afghanistan are “priority one countries,” and the volume of coverage of Pakistani media has more than tripled since 2003 as greater monitoring resources were brought to bear on that region.³¹

Unlike FBIS, whose budget fell under the secrecy guidelines of

the intelligence community that housed it, BBC publishes basic annual financial figures, offering some insights into the scope of its operations. During 2008/2009, its total budget was approximately £28.7 million (\$45.9 million), of which £24.6 million came from the British government, £1.4 million from commercial licensing, and £2.6 million from lessees, interest, and income from the Open Source Center. Expenditures included £15.1 million for staff, £3.6 million for “accommodation, services, communications, maintenance, and IT,” £479,000 for copyright clearances, £3.8 million for “other” and £3 million for depreciation.³² The governmental portion of its funding

for 1994/95 was approximately £18.4 million (\$28.7 million), suggesting generally stable levels of governmental support over the past decade and a half.³³

Editorial Process

FBIS and SWB are renowned for the extremely high quality of their translations, which often capture the tone and nuance of the original vernacular. Such translation quality requires a high level of editorial input, including iterative revision processes in both services. Changes in translation, however, manifest themselves in ways that complicate con-

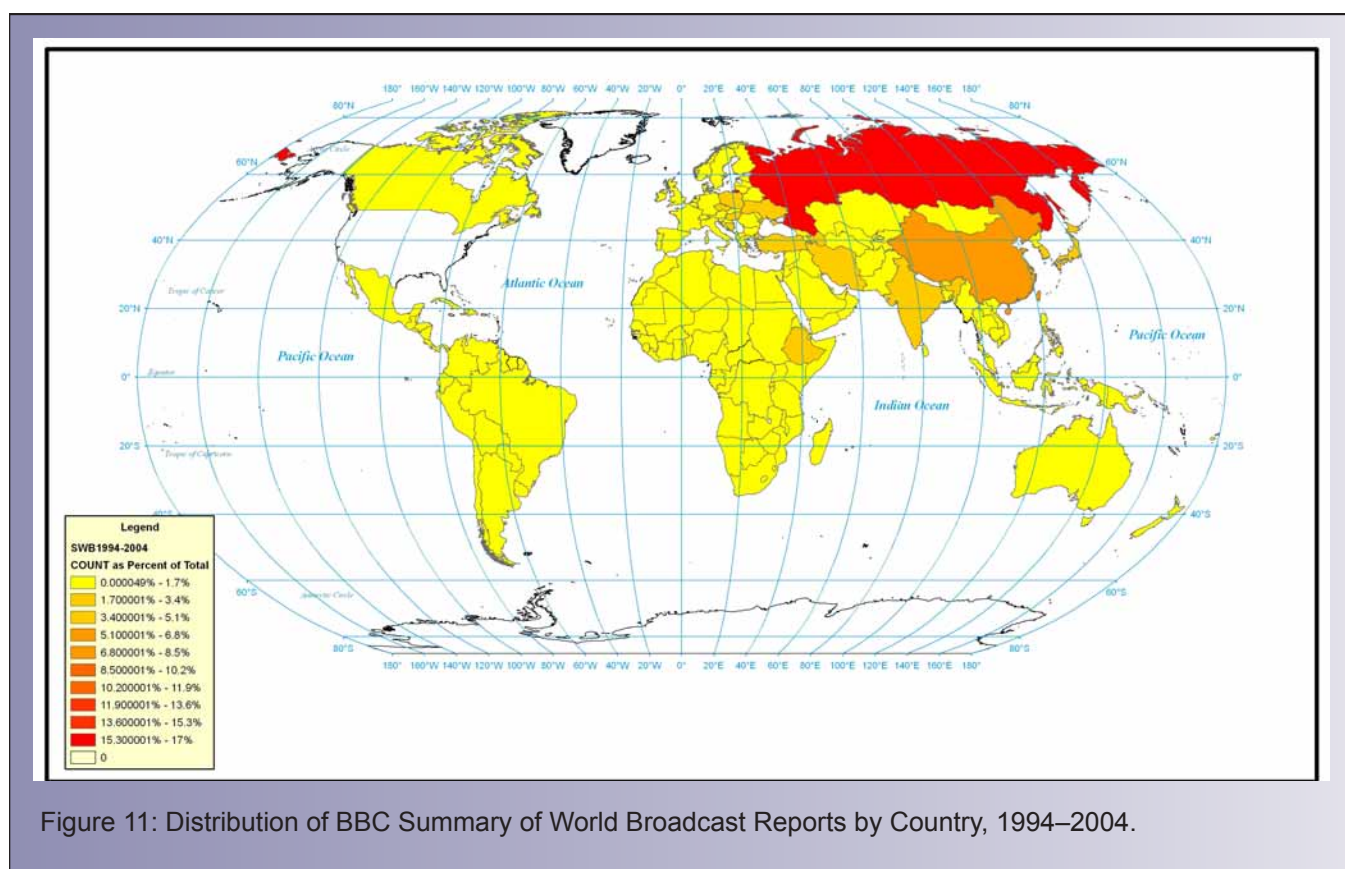


Figure 11: Distribution of BBC Summary of World Broadcast Reports by Country, 1994–2004.

tent analysis of the FBIS and SWB databases.

In FBIS it is possible that an editor or a downstream consumer might inquire about aspects of a given translation for clarification or amplification and prompt a retranslation. This is especially prevalent with broadcast transmissions, which can suffer from interference that make passages unclear.

But FBIS methods for accounting for such changes were inconsistent. An FBIS translation or transcription that was substantially changed might have been reissued to the wire. In some cases a notation

was provided, such as a 1998 FBIS report drawn from Radio France Internationale that noted at the beginning: “Corrected version of item originally filed as ab0909100698;³⁴ editorial notes within body of item explained changes made.” The corrected report was assigned its own unique FBIS ID, AB0909113898, and since no structured field existed in the database system on the CDs to connect the two reports, an analyst would have to read the note in order to recognize that reports are the same item. Researchers conducting automated queries, such as a time-series analysis, would find this item double counted.

Unfortunately, acknowledgment of revisions in both collections is the exception rather than the norm. The FBIS reports studied show duplication of about 1 to 2 percent per day. In some cases, it is only the title that changes or a duplicate report may simply have been an error, such as a 5,530-word report from 2001 that was reissued later the same day without the last 731 words.³⁵ In another case, a 1 January 2001 article about NATO changed “Foreign Minister” Colin Powell to “Secretary of State” and the fate of the “enlargement” of the North Atlantic Alliance became simply the fate of the “Alliance itself.”³⁶ A sentence was also moved down in the

first paragraph, together with several other smaller changes, altering nearly 10 percent of the total text. In both cases, the duplicate reports had their own unique identifiers but contain no information linking them to their originals.

For the entire period 1979–2008, the Lexis SWB archive contains 4,694,122 reports (discounting separate summary reports of fuller accounts). Analysis of the reports showed that nearly 1 million of these reports were duplicates.

SWB content accessed through Lexis for the years 1998–2002 showcases this revision process and underscores the challenges for content analysts. Curiously, explanations for this duplication differ over two periods of time over these five years. The easier period to explain is the period from March 2001–December 2002, when nearly half of all reporting was duplicated. Duplicates during this period are in most instances identical copies of earlier reports, with the exception of some extraneous formatting characters. Simple textual comparison of all reports issued on each day identified the duplicates. This accounted for about 700,000 duplicates.

The remaining reports, which run from January 1998 through March 2001, present a much more significant analytical challenge. The duplicates during this period are not identical copies. They are retranslations of earlier reports. Some only have changes in titles, for example, “inaugurated” becoming “set up” or “Montenegrin

outgoing president” changing to “outgoing president.”³⁴ However, most include changes to the body text itself, such as a 24 January 1998 Romanian Radio broadcast that first appeared in Lexis on the 25th, with a revised edition issued the following day.³⁵ Seven changes were made to the body text, including “make” changed to “do” and “make the reform” becoming “carry out reforms.” Several words were changed from singular to plural or vice-versa, while monitor’s comments were inserted to indicate the speaker for different passages. In all, nearly 4 percent of the report’s total text was changed.

Linking articles containing multiple substantive changes of this kind is a non-trivial task: sentence order may be revised, words changed, and phrases added or deleted. Simple textual comparison will not suffice and more advanced detection tools are required. Titles can also change. Unfortunately, SWB uses the same timestamp in the source citations of all reports from the same broadcast, meaning that header fields do not provide information to help distinguish duplicates. Instead, full text document clustering is required, a technique that computes overlap in word usage between every possible combination of documents for a given day. If two documents overlap by 90 percent or more, they are considered duplicates.

Such an approach allows for fully automated detection and removal of duplicates, with extremely high accuracy (a ran-

dom sample of days checked, for example, revealed no false positives). In all, the 38 months of this period exhibit an average of 42-percent duplication, with a high of nearly 65 percent in January 2001. With clustered duplicates removed, a total of 3,700,761 unique reports remain from the original nearly 4.7 million reports.

Even this approach can only identify reports with relatively minor alterations. Wholesale rewrites—those that keep factual information the same, but substantially or completely altered wording—cannot readily be detected through purely automated means. For example, a January 1998 report about rice prices was initially released containing numerous monitor comments indicating unclear transcription. The 93-word transcript was rereleased nine days later as a 50-word paraphrased edition.³⁶ A 303-word transcript the same month concerning enactment of a tax law in Russia was rereleased six days later, cut nearly in half, again with heavy paraphrasing and rewriting.³⁷ In both cases the “Text of Report” header denoting a full-text transcript was removed from the subsequent report, suggesting an explicit decision on the part of the monitoring staff to switch from a literal translation to a paraphrased summary. A manual review of content during this period suggests that this activity may be restricted to broadcast content, which presents the greatest challenges for accurate transcription.

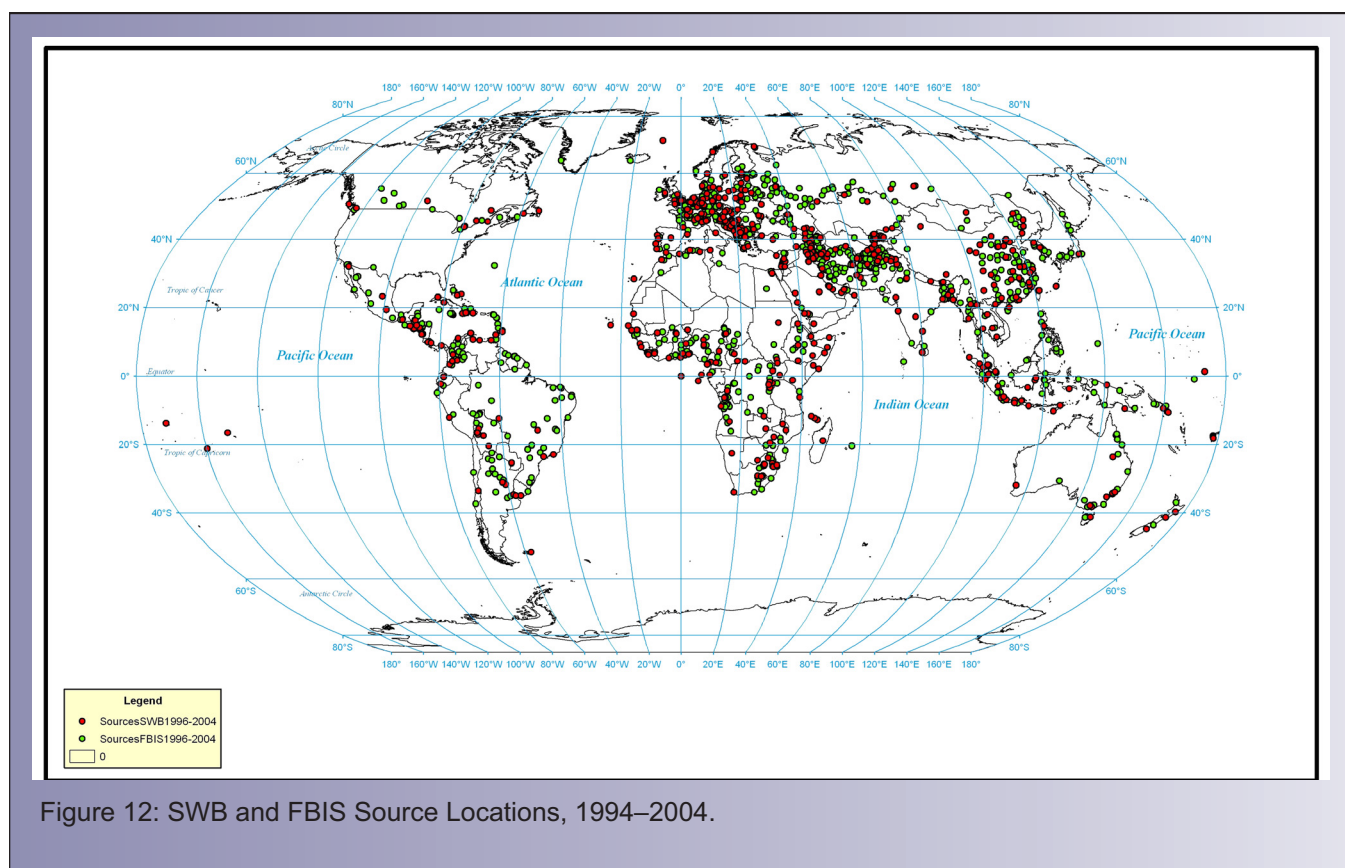


Figure 12: SWB and FBIS Source Locations, 1994–2004.

SWB and FBIS Coverage Compared

FBIS and SWB had a long history of sharing content. The maps on this and the next page (figures 12 and 13) show the similarity of the two services's geographic emphases. (Their Pearson correlation is $r=0.84$ [$N=191$], suggesting very strong overlap.)

Unfortunately, source references are constructed very differently in the two collections, so it is only possible to compare source listings geographically. Figure 12 locates all SWB and FBIS sources during this period. To simplify the map rendering, if SWB and FBIS both have a source at a given location, the FBIS map point may be obscured by the SWB point.

The data show that FBIS draws from a larger selection of sources in a broader geographic range than does SWB.

Unlike FBIS, SWB draws some content from sources based in the United States (primarily US sources aimed at foreign audiences), but those account for only a small fraction of its content and are not shown here. FBIS is a much-higher-volume service, generating an average monthly volume of just over two and a half times that of SWB from 1993–2004, which may also account for the larger number of sources.

Shifting Coverage Trends

Because SWB content is available in digital format back to 1 January 1979, it is possible to analyze a 30-year span to trace

the evolution of geographic coverage of monitored material.

As shown in figure 13, which illustrates the total change in coverage density from 1979 to 2009, relatively large increases have taken place in coverage of Iran and Pakistan; little change can be seen in other Middle Eastern nations, notwithstanding increased Western military presences in Iraq and Afghanistan; and declines have occurred in coverage of Russia and China, where the decline has been the most pronounced. If SWB coverage can, indeed, be used to infer levels of US coverage of open sources today, these data support the argument that open source resources are not, by and large, retasked to military conflict zones and provide instead a strategic resource.

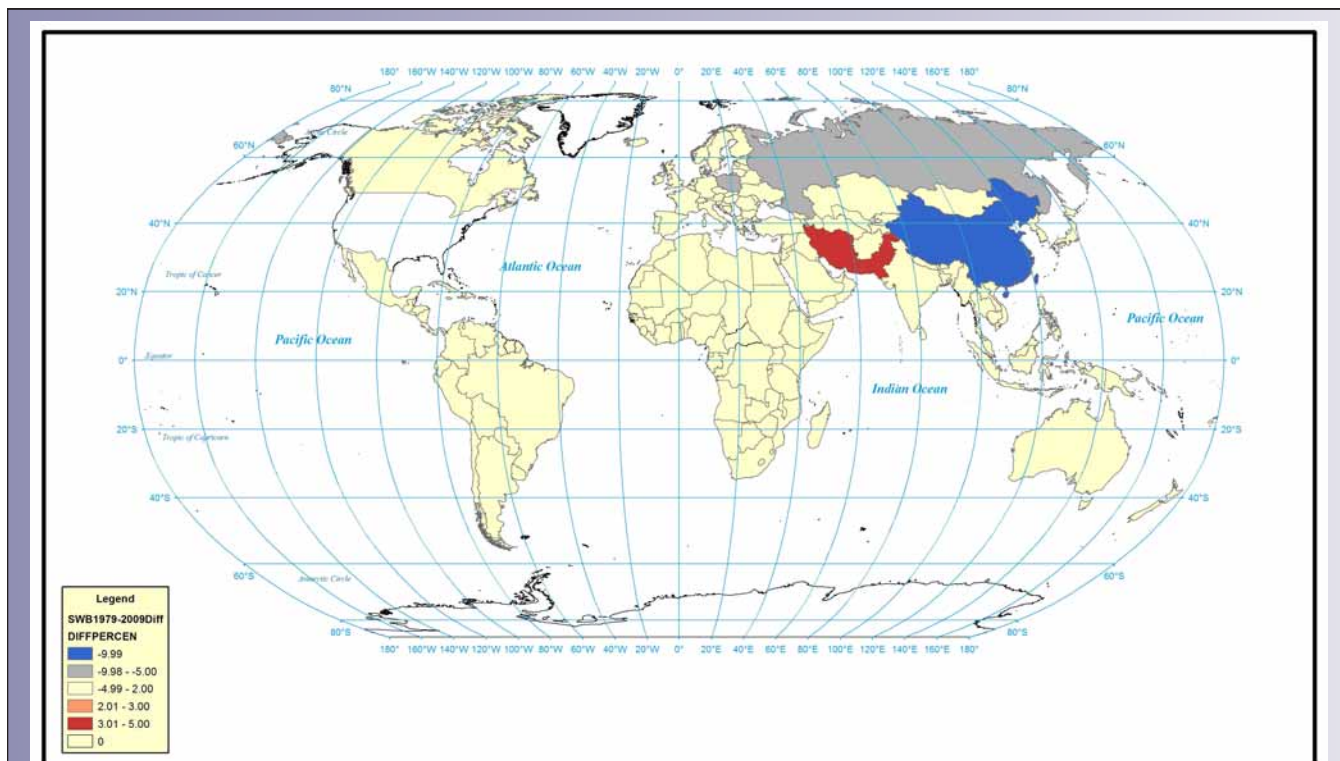


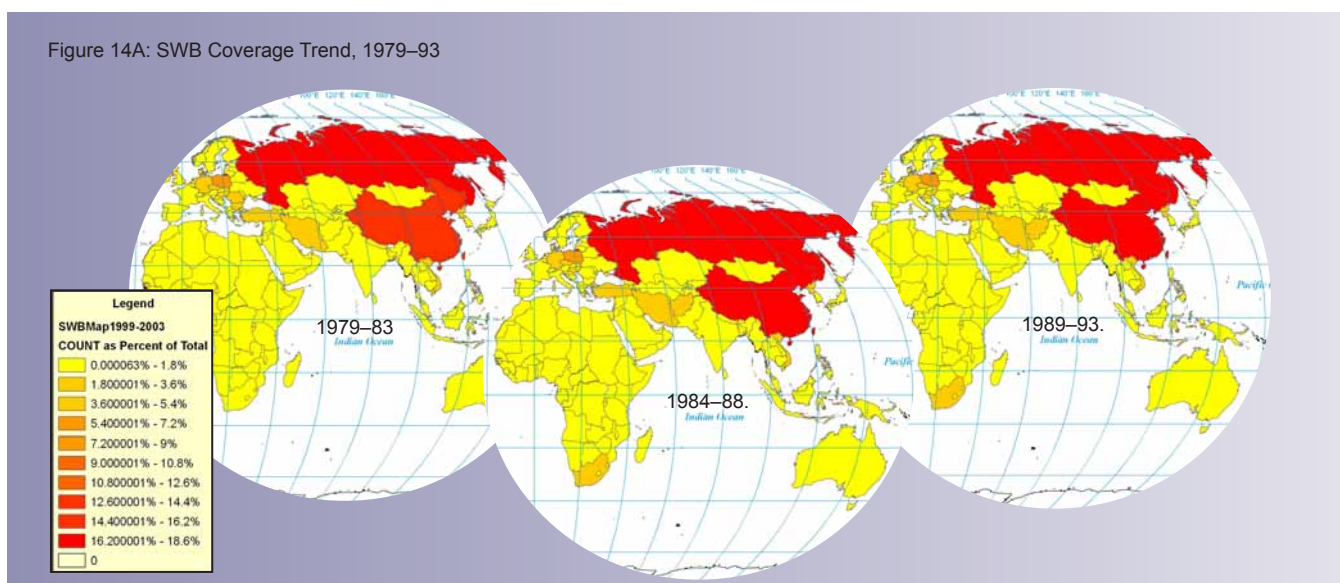
Figure 13: SWB Coverage Density Change, 1979–2008.

Figures 14A and 14B show coverage shifts in five-year increments during this period. (Western Hemisphere countries are not shown because there was relatively little change in the period.) These graphs further highlight the

evenness of SWB coverage throughout the world and the sustained emphasis on Russia and China, mirroring FBIS's focus on these two countries. The impact on analysis of such stable sourcing cannot be overstated. While countless studies

examine the geographic biases in Western reporting of international events, SWB appears to be largely immune to such selection biases, with African and Latin American countries receiving nearly the same

Figure 14A: SWB Coverage Trend, 1979–93



attention as their European counterparts.

The relatively intense coverage of Russia and China, however, is more troubling for those seeking to do broad-based research. All six maps use the same color scale, showing that Russian emphasis has remained nearly constant for three decades. Emphasis on China, on the other hand, has decreased nearly linearly over this period.

Increases in coverage of some areas evident in these maps—Greece, Poland, and India, for example—track with heightened security concerns during the periods.

Conclusion

Notwithstanding recent criticism of US neglect of open source intelligence, the record of US and British collection of such intelligence evident in publicly available collections reflects a longstanding US and British

US and British OSINT services' ability to penetrate into the non-Western world will make their products central to the next wave of social science research.

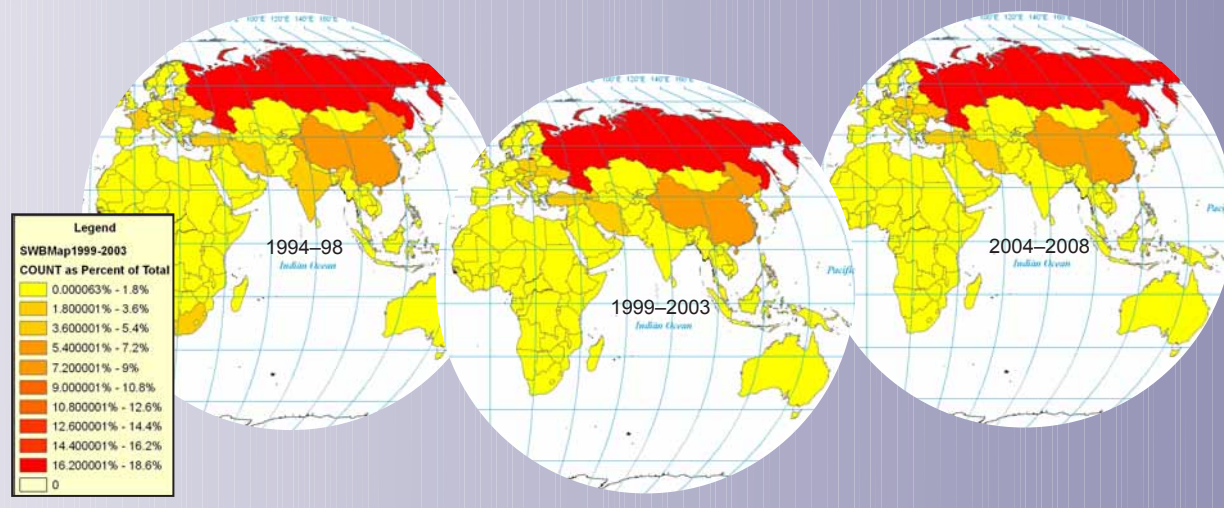
understanding of the importance of realtime, uniform monitoring of the media output of nations around the globe.

For the academic researcher, the two services in effect act as time machines, allowing social and political scientists, historians, and others to turn back the clock to revisit events in innovative ways. While the goals of intelligence analysts using OSINT are different from those of academic researchers, their needs and methodologies are similar. On the academic side, content analysts of international events have historically been limited by the constraints of commercial news databases dominated by Western media. With increasing globalization of so many social, economic, and political phenomena, scholars will have to abandon reliance on Western newspapers and look elsewhere.

The ability of US and British OSINT services to penetrate into the non-Western world will make their products central to the next wave of social science research. They operate as an almost ideal strategic monitoring resource, with nearly even coverage across the globe, and offer a unique view into the broadcast news media that dominate many regions of the world. Their political and economic focus and full-text English translations make them a powerful resource for international news studies. As the world grows smaller, OSINT offers academic scholars an unparalleled complement to existing commercial databases and provides a unique opportunity for academia and government to collaborate in furthering our understanding of the global news media and the insights it can provide into the functioning of societies.



Figure 14B: SWB Coverage Trend, 1994–2008



Endnotes

1. Kalev Leetaru, "An Open Source Study of International Media Coverage of the WorldCom Scandal," *Journal of International Communication*, 14, no. 2.
2. George Gerbner and George Marvanyi, "The Many Worlds of the World's Press," *Journal of Communication*, 27, no. 1 (1977): 52-66.
3. Ben Barber, "CIA Media Translations May Be Cut: Users Rush to Save Valuable Resource," *Washington Times*, 30 December 1996.
4. Wesley Pruden, "The Spies go AWOL when Kabul falls," *Washington Times*, 27 November 2001.
5. Hadley Cantril, *The Human Dimension: Experiences in Policy Research* (Rutgers, NJ: Rutgers University Press, 1967).
6. Princeton Listening Center, Online Records, 1939–1941, "Finding Aid," <http://diglib.princeton.edu/ead/getEad?id=ark:/88435/0v838057k> (accessed 6 April 2009).
7. U.S. National Archives and Records Administration, *Records of the Central Intelligence Agency [CIA]: Records of the Foreign Broadcast Information Service and its Predecessors (1941-1974)*, Record Series 263.3. Abstract available online at <http://www.archives.gov/research/guide-fed-records/groups/263.html?template=print#263.3>; Niles J. Riddel, "Opening Remarks, First International Symposium on National Security and National Competitiveness: Open Source Solutions," 2 December 1992, available online at <http://ftp.fas.org/irp/fbis/riddel.html> (accessed 1 July 2007); U.S. National Archives and Records Administration, *Records of the Foreign Broadcast Intelligence Service, Civilian Agency Records, Other Agency Records* RG 262, <http://www.archives.gov/research/holocaust/finding-aid/civilian/rg-262.html> (accessed 1 December 2007).
8. Cantril, *Human Dimensions*, 30.
9. Stephen Mercado, "FBIS Against the Axis, 1941–1945: Open-Source Intelligence From the Airwaves," *Studies in Intelligence* (Fall-Winter 2001), available online at https://www.cia.gov/library/center-for-the-study-of-intelligence/sci-publications/csi-studies/fall_winter-2001/article04.html (accessed 1 December 2007). See also: Foreign Broadcast Information Service. *FBIS in Retrospect*, (Washington, DC: 1971).
10. Riddel, "Opening Remarks."
11. Brian Rotheray, *A History of BBC Monitoring*, http://www.monitor.bbc.co.uk/about_us/BBCMhistory%20revisions%20x.pdf (accessed 6 November 2009).
12. Michael Goodman, "British Intelligence and the British Broadcasting Corporation: A Snapshot of a Happy Marriage," in *Spinning Intelligence, Why Intelligence Needs the Media, Why the Media Needs Intelligence*, Robert Dover and Michael Goodman, eds. (New York: Columbia University Press, 2009), 117–32.
13. Joseph E. Roop, *Foreign Broadcast Information Service, History, Part I: 1941–1947*, available online at <https://www.cia.gov/library/center-for-the-study-of-intelligence/csi-publications/books-and-monographs/foreign-broadcast-information-service/index.html> (declassified, 2009); Brian Rotheray, *New Risks of Crisis-Fresh Perspectives from Open Source*, 2001, slides, http://www.earth-intelligence.net/dynama-ster/file_archive/040319/43bf89d1718a95df6f78847b3c4aad0b/OSS2001-X1-18.pdf (accessed 7 May 2009).
14. Rotheray, *New Risks*.
15. Riddel, "Opening Remarks;" William Studeman, "Teaching the Giant to Dance: Contradictions and Opportunities in Open Source Within the Intelligence Community," presented at First International Symposium on Open Source Solutions. Available online at <http://www.fas.org/irp/fbis/studem.html>.
16. Riddel, "Opening Remarks."
17. Ibid.
18. Studeman, "Teaching the Giant to Dance."

19. Ibid.
20. Natalie Gawdiak, (1997, January 3). Internet Posting to <http://www.fas.org/irp/fbis/guestbook.html>, 3 January 1997 (accessed 11 November 2009).
21. Andrew Reeves, Stephen Shellman, and Brandon Stewart, "Fair & Balanced or Fit to Print? The Effects of Media Sources on Statistical Inferences," presented at the International Studies Association conference, 22–25 March 2006, San Diego, CA.
22. Robert L. Solso, "The KAL 007 Tragedy: The Use of FBIS in Cross Cultural Research," *Journal of Government Information*, 13, no. 2 (1986): 203–8. In reviewing FBIS content, Solso found that its coverage of Russia corresponded very closely with the publication *Current Digest of the Soviet Press*.
23. David Ensor, "Open source intelligence center," CNN.com, 8 November 2005, available online at <http://www.cnn.com/2005/POLITICS/11/08/sr.tues/> (accessed 1 July 2007).
24. Bill Gertz, "CIA Mines 'Rich' Content from Blogs, *Washington Times*, 19 April 2006; Robert K. Ackerman, "Intelligence Center Mines Open Sources," http://www.afcea.org/signal/articles/templates/SIGNAL_Article_Template.asp?articleid+1102&zoneid=31 (accessed 1 July 2007).
25. Riddel, "Opening Remarks."
26. Ibid.
27. People's Radio Network, 8 August 1996, 2300 GMT, as reported in FBIS; Radio France International, 12 August 1996, 1230GMT, as reported in FBIS.
28. William A. Rugh, *Arab Mass Media: Newspapers, Radio, and Television in Arab Politics*, (Westport, CT: Praeger, 2004).
29. Rotheray, *New Risks of Crisis*.
30. BBC Monitoring, *Factiva Content Community*, Case Study, 2001, <http://www.factiva.com/content-comm/casestudies/bbccasestudy.pdf> (accessed 6 November 2009).
31. BBC Monitoring, *Annual Review 2008/ 2009*, http://www.monitor.bbc.co.uk/images/BBCM_review_08_09_LR.pdf (accessed 6 November 2009).
32. Ibid.
33. Rotheray, *New Risks of Crisis*.
34. "Congo-Kinshasha: DR Congo Rebels React to Victoria Falls Summit," 9 September 1998, as received by FBIS 0730GMT from Paris Radio France Internationale in French.
35. "Senior Cleric Discusses Religious Pluralism," 1 January 2001, as translated by FBIS from Qom Pasdare Eslam in Persian, 8–13.
36. "Uncertain Fate of NATO Enlargement, Slovakia's Integration Examined," 1 January 2001, as translated by FBIS from Bratislava Pravda in Slovak, 6.
37. "Special hotline inaugurated between US, Israeli defence ministers," 29 January 1998, as received 1800GMT by SWB from Channel 2 TV, Jerusalem, in Hebrew; "Problems are not solved with tear gas—Montenegrin outgoing president," 14 January 1998, as received 1800GMT by SWB from Serbian Radio, Belgrade.
38. "No political crisis in Romania, president says," 24 January 1998, as received 1700GMT by SWB from Romanian Radio, Bucharest, in Romanian.
39. "Agriculture official urges calm over rice price," 12 January 1998, as received 0900GMT by SWB from Radio Television Malaysia Radio 4, Kuala Lumpur, in English.
40. "Yeltsin signs law on income tax," 2 January 1998, as received 1243GMT by SWB from ITAR-TASS World Service, Moscow, in English.